Efficient Votes Storage in a Non-Interactive Dining Cryptographers (NIDC) Protocol

Pablo Garcia[1], Silvia Bast[1], Germán Montejano[1,2]

1. FCEyN (UNLPam). Uruguay 151, Santa Rosa, La Pampa, Argentina
{pablogarcia, silviabast}@exactas.unlpam.edu.ar.
2. FCFMyN – (UNSL). Ejército de Los Andes 950, San Luis, Argentina
gmonte@unsl.edu.ar

Abstract - This paper shows the behavior of a storage technique for anonymous data, based on parallel channels implementation. At the beginning, it was conceived to apply on electronic vote. However, it may be generalized to whatever situation that requires anonymity and demands very high level of security respect from loss of information. Furthermore, a formula mistake that was slipped in previous publications has been corrected here. Finally, results of simulations are shown, to analyze the behavior of the equations proposed.

Key words: E-Voting - NIDC – Anonymity – Parallel Channels.

1. Introduction

There are many real-world problems that require very high levels of security with respect to information use. A good example of an application, whose level of demand in that sense is highest, is the electronic voting. In such an application, the condition of anonymity of a voter indefinitely is crucial. But indisputably it must be ensured that the result of the ballot accurately reflects the will of the electorate.

It is obvious that the interests involved are very transcendent and that, therefore, attempts to perform a fraud may occur with high probability. And if consider in detail, many aspects of the process, whose safety must be ensured, appear. In this case, the behavior of an alternative proposal for the storage of votes, based in the implementation of parallel channels of slots will be analyzed.

The results shown in this paper belong to a research line that began in 2013 [1]. Within this scope, it is searched to define the exact assurance level requested for anonymity in an electronic voting scheme. In [2] it is concluded that it is necessary to give unconditional security for the privacy, because it must be protected indefinitely. Otherwise, votes must be kept for a finite period of time.

Consequently, protocols that verify that condition reach most importance. Particularly one of the most interesting is Dinning Cryptographers, which is described in detail in [3]. This protocol is resourceful and it covers the requirement to guarantee unconditional privacy.

The model may be described as follows:

"Three cryptographers share a dinner in a restaurant. When the time to pay comes, the waiter tells them that the addition has already been paid and that who did

it, does not want that his identity is revealed. Cryptographers want to know if any of the guests was the one who made the payment, or if it was paid by someone external to the group of diners. They only want to know whether any of them paid or not."

Raised in this way, the solution found is:

"Each of the diners throws a coin. He looks at the result and shares with its neighbor on the left. Then, each of them looks exactly two currencies, self and neighbor who shares with him. Finally, each one should indicate whether the two currencies that could be observed are "equal" or " different" with the condition that if any of them paid the addition, he should lie about his statement."

In the conditions described, if the number of cryptographers proclaiming "different" is odd, the payer is in the group of diners. An even number, otherwise, indicates that the payer is external to the group.

Dining cryptographers presents unconditional security levels, with regard to anonymity associated with the issuance of certain information, through public channels. The initial problem is based on three participants exchanging only one bit of information, but it generalizes to any number of participants and any volume of information naturally and without significant complications.

It is necessary to highlight some points:

1) It is considered that the currencies used provide a truly random experiment result with respect to "throw the coin", so that, Pr (*"Heads"*) = Pr (*"Tails"*) = ½.

2) This scheme gives correct results only if it has the honesty in the response of all participants. If someone paid but he does not lie in his statement the model does not guarantee right conclusions. The same applies if a diner, who did not pay the bill, does not tell the truth when comparing both currencies.

3) This scheme works correctly if a single payment is made for dinner. If the waiter had accepted two or more anonymous payments, the conclusions will not be correct. This point relates to the above: the original model works properly only if all participants show an honest behavior.

4) If the required conditions are right, none of the diners get any information on the identity of the payer. As stated, if the payer is external, anonymity is assured. If the payer belongs to the group, it is easy to analyze cases to conclude that a cryptographer who did not pay the bill does not receive any information that allows deducing the identity of the payer. This property makes it very attractive scheme, since the central goal (anonymity) is obtained by default, as included in the scheme without additional effort.

5) One element of great value in Dining Cryptographers is the unconditional security given to the anonymity. A system which allows transmission of messages ensuring the unconditional anonymity of the source is provided. For any information exchange scheme in which privacy is desired, dining cryptographers becomes a very attractive scheme.

To understand the underlying reason why anonymity is ensured, it is necessary to put in place a diner who has not paid the bill (*C*). Obviously, the case in which the payer is external guarantees anonymity based on the assumptions of the scheme. Therefore, one must analyze the case where the payer belongs to the group of cryptographers. This can only have two alternatives:

- *The two currencies that C looks are equal.* In this case, necessarily, one of the other participants indicated "equal" and the other "different".

Therefore, if the currency that he could not observe were equal to the one he saw, the participant who said "different" is the payer. And, conversely, if the unknown currency was different from which he could see, the payer is who expressed "equal". However, both states are equally probable in the scheme. Therefore, it is not possible to obtain information related to the payer.

- *The two currencies that C looks are different.* In this case, it is inevitable that the other two cryptographers express coincident results. If both say "different", the payer will be the participant who is closest to the currency matching the result of the hidden coin. And if both proclaim "equal", the payer is who is closer to the currency that differs from the value of the hidden coin. As in the previous case both situations are equally probable. Consequently, any information that betrays the identity of the payer may not be obtained.

To complete the analysis, the concept of view, which will allow us to demonstrate the safety of the proposed scheme, is defined:

A view is a random variable describing what set of information has a particular participant when the process is finished. For example, at the end of the election act if it is a scheme of electronic voting.

If we may prove that the view obtained by any user may not determine the choices made by the other participants in any case, anonymity is ensured. In particular, to analyze the scheme Dining Cryptographers, we may distinguish the following elements and analyze which ones are visible for each participant:

- Coins: according to the mechanics of the DC model, each participant sees its own currency and its neighbor's on the left. We denominate $x_i \in \{Tails, Heads\}$ to the value obtained in the action of the coin toss i.
- Investment information: The value of this element will be $m_i \in \{True, False\}$. If m_i is True, it implies that the participant paid the bill and, therefore, lies about the obtained result when comparing the two currencies that can be observed. A value False implies the opposite.
- Information comparing two currencies: For this data $r_i \in \{Equal, Different\}$ will be used. Obviously, a value Equal implies that the participant i declares that the values of the two currencies which he may see are coincident.

Therefore, in these terms, initial security scheme may be observed. It is sufficient to analyze the views that a particular participant has for all possible cases, since the symmetry of the scheme ensures that the conclusions may be generalized for all participants.

The views that the participant A_1 could have available is then analyzed. The cases are:

- The payer is external. In this case, anonymity is guaranteed.
- *A1* is the payer. This situation is also trivial.
- *A2* is the payer. The view that A1 has available is the following:

$$V1 = (x_1, -, x_3, m_1, -, -, r_1, r_2, r_3) \quad (1)$$

- A_3 is the payer. In this situation, A_1 observes:

$$V_1 = (x_1, -, x_3, m_1, -, -, r_1, r_2, r_3) \quad (2)$$

It is clear that the views of the last two cases are coincident. Specifically, the values of x_1, x_3, m_1 and r_1 have the same probability distribution in both cases; r_2 and r_3, instead, present opposing values depending on who paid the bill. However, that does not give additional information to A_1, because such values depend on x_2, value he does not know and has the same probability to take either of two possible values.

Therefore, if any of his colleagues paid dinner, A_1 cannot tell who it was, because:

$$\Pr(A_1) = \Pr(A_2) = ½ \quad (3)$$

2. Non - Interactive Dining Cryptographers (NIDC)

The analysis is focused on a derivative one called Non Interactive Dining Cryptographers (NIDC, [4]), that relaxes the condition of concurrency online for all participants, condition that is present in multiple real-world problems.

The idea is that, through the use of blind signatures introduced by Chaum [5], the voter obtains a valid vote of the authorities of the process election. In particular, the protocol presented in [6] could be used. This protocol allows the voter communicate with authorities to send a blind vote. They respond by signing (blindly) the vote and resending it to the voter. It should be noted that the process is perfect and that all options are mathematically equal, the reason explaining why authorities cannot deduct any information related to the voter's option.

Obviously, the authorities should record each vote, so that no voter may cast vote several times. Similarly, both parties must sign their messages and keep records of them for the purpose of solving any subsequent dispute.

Upon receiving the message from the authorities, the voter retrieves it and he can verify that it contains a valid and signed vote.

NIDC uses a storage model based on a single vector of slots. The anonymity is guaranteed by random position where a vote is stored. The fact of randomness brings the outcome of collisions. A collision occurs when two or more votes are stored in the same slot. That results in the loss of the coincident votes. In this context, the proposed model in Figure 1 may be explained by Birthday Paradox [7] which states:

"In a group of 23 people, the probability that there are at least 2 who share the same birthday is very close to ½."

This assertion is little auspicious for the purposes of this research. Associated vector size is relatively large respect to the sample; however, the associated security level is far from what could be acceptable in practical applications. An E-voting system in which the probability of loosing at least one vote is next to ½ lacks importance.

Graphically, the original scheme is shown in Figure 1.

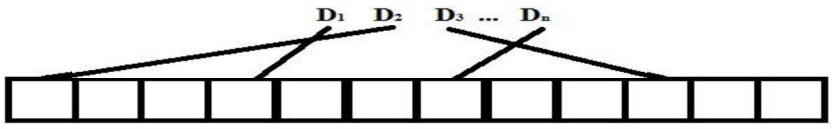

Fig1: Original Storage Scheme NIDC

The tendency of behavior of a scheme like Birthday paradox keeps little interesting values whatever the parameters were. The amount of positions that must maintain the vector to obtain acceptable security levels is very significant. For example in the typical case (365 days and 23 people) the probability of a collision approaches ½ although, there exists 342 dates where nobody was born.

It is possible to think about a scheme that improves this redundancy in an efficient way. This alternative is described in the next section.

3. NEW PROPOSED STORAGE SCHEME TO NIDC

According to what is stated in the previous section, it becomes of great interest to analyze deeply the possibility to find alternative methods that improve the use of storage, since the single vector scheme requires significant number of positions in order to ensure an appropriate security level. In [8] an alternative proposal is exposed. It is shown in Figure 2.

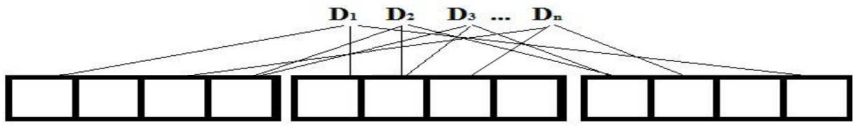

Fig 2: New Alternative For Storage.

In [9] another approach is shown, related to NIDC. It is based on parallel or serial channels. Conversely, this document goes forward to what was exposed in [10], where a series of equations are exposed. Those equations describe behavior of the model, based on the following parameters:

T: #total slots to implement.
S: #slots on each channel.
Q: #parallel channels to implement.
Q_t: #parallel channels to implement. (Theoretical).
Q_p: #parallel channels to implement. (Practical).
N: # voters.
V_i: i_th vote.
R_{ij}: event which indicates that V_i occupies j_th slot.
C_{ijk}: event that occurs when V_i collides with V_j in channel k.
B_{ij}: event when V_i loses in cannel j.
A_i: event which indicates that at least a multiple collision is produced.
X: event which indicates that no vote is lost simultaneously in all channels.
L: #votes that are lost in all channels simultaneously.

I: event which indicates that at least a multiple collision is produced.

The equations exposed in [10] are the following:

For a fixed number of voters *N*, the recommended number of slots (*S*) for each parallel channel is given by the formula:

$$S = \left\lfloor \frac{N}{\ln 2} \right\rfloor + 1 \tag{4}$$

For given values of T and N, there exist an optimal number of parallel channels. Such value is expressed by:

$$Q_t = \ln 2 \frac{T}{N} \tag{5}$$

That formula should be the next integer.

$$Qp = | Qt | + 1 \tag{6}$$

As explained previously, it is rounded to the next integer, since the value of Q applied must be necessarily integer.

The expected value for the *PLV* variable (Percentage of Lost Votes) is obtained by applying equation:

$$|PLV| = (1 - e^{-\frac{N}{S}})^Q \tag{7}$$

An appropriate lower bound for the probability of "no votes are lost" is obtained by computing equation:

$$\Pr(X) > 1 - (\tfrac{1}{S}(N-1)^Q)^N \tag{8}$$

This last equation corrects the one published in [8] where an error was slipped. Indeed, the formula is correctly developed, as follows:

For $Q = 1$:

If V1 drops in slot 1, the probability to collide with V_2, consists on that both drop in slot 1:

$$\Pr(C_{121} | R_{11}) = \frac{1}{S}\frac{1}{S} = \frac{1}{S^2} \tag{9}$$

Then, the probability that votes V_1 and V_2 collide in whatever slot is:

$$\Pr(C_{121}) = \frac{1}{S^2} S = \frac{1}{S} \tag{10}$$

$$\Pr(C_{121}) = \Pr(C_{1j1}) \; \forall j \in \{3..N\} \tag{11}$$

Let *I* = "Multiple Collisions occur", where, "Multiple Collision" means that three or more votes are stored in the same slot. Consequently, the probability that V_1 is lost in the only channel is given by the expression:

$$\Pr(B_{11}) = \frac{1}{S}(N-1) - \Pr(I) \tag{12}$$

From the previous equation, another one, which is more appropriate is derived. This equation is based on the probability V_1 is not lost.

$$Pr(\overline{B\ }) = 1 - \frac{}{S}(N-1) + Pr(I) \tag{13}$$

The value for *Pr(I)* is low but positive, then:

$$Pr(\overline{B\ }) > 1 - \frac{}{S}(N-1) \tag{14}$$

The previous formula lets explain to a voter the probability that his vote is lost or not in a single channel scheme. The accuracy of the formula increases when $(S \to \infty) \wedge (N \to \infty)$.

When more channels are added, it can be asserted:

$$Pr(A_i) = Pr(B_i\ )^Q > \left(\frac{}{S}(N-1)\right)^Q \ \forall i \in \{1 \dots N\} \tag{15}$$

Finally, an appropriate lower bound for the probability of *X="no vote is lost simultaneously in all channels"*, is obtained applying the following equation:

$$Pr(X) = Pr(\overline{A\ }) \cap Pr(\overline{A_2}) \cap \dots \cap Pr(\overline{A_N}) \tag{16}$$

But:

$$Pr(\overline{A\ }) = Pr(\overline{A_2}) = \cdots Pr\ (\overline{A_N}) \tag{17}$$

Moreover, they are self independent events, then:

$$Pr(X) = Pr(\overline{A\ })^N \tag{18}$$

Or what is the same:

$$Pr(X) = 1 - Pr(A\ )^N \tag{19}$$

Finally, it can be asserted:

$$Pr(X) > 1 - ((\frac{}{S}(N-1))^Q)^N \tag{20}$$

Precisely, at this point, the formula published in [8] has an error, consisting on the location of a parenthesis. The formula described above (20) is the correct form.

This lower bound is very useful to describe the probability of losing votes in a real voting with mentioned parameters. Definitely, it allows knowing previously the exact probability of no losing votes during the process.

4. SIMULATIONS

Given formulas above and knowing about error in formula ([8]), a simulator has been implemented which have two main aims:

1) To verify the correctness of formulas.
2) To bear out that the approach of storing in parallel channels optimizes the results in terms of several variables which can be considered.

The simulator is implemented allowing the following inputs:

1) Total number of slots to implement (T).
2) Number of voters (N).
3) Quantity of parallel channels to implement (Q).
4) Quantity of election acts that will be simulated by session (R).

The simulator verifies that the total number of slots (T) is a multiple of quantity of parallel channel, because the quantity of slots in each channel (S) must be an integer number.

When the simulation is complete, the following information can be obtained:

1) Total of successful votes (SV).
2) Total of lost votes (LV).
3) Quantity of runs where at least one vote is lost (R)
4) Quantity of runs (Votings) without lost votes ($RWLV$)
5) Quantity of runs (Votings) with lost votes (RLV)
6) Best case, that is to say, how many votes were lost in the most successful run (BC).
7) Worst case, that is to say, how many votes were lost in the less successful run (WC).

For all runs the following values for the parameters were selected:

T = 480 slots.	N = 120 voters.	R = 1.000.000

The successive simulations were executed with Q = 1..5. Q_t, using the appropriate formula, gives 2,772588722 channels. Consequently, it is expected that the best values will be obtained using Q_p = 3.

A. Verifying correctness of the formulas

Multiple simulations were executed to bear out the correctness of formulas (5) y (6). As it was explained previously, (5) indicates optimum theoretical value for Q_t y (6) the next integer. In all simulations the best results were obtained dividing t slots in Q_p parallel channels. For the purpose of illustrating the situation, figures 3, 4 are examples of the results obtained with selected values of T y N.

Hundreds of simulations were executed and in all cases optimum results occurs when Q_p is used.

Figure 3 shows percentage of lost votes for Q = 1..5. The lower loss occurs when Q=3.

Figure 4, shows quantity of successful votes. Also, in this case the best result occurs when Q=3.

B. Behavior of parallel channels technique.

At this point there are several facts that validate the method.

1) When 1.000.000 runs are fulfilled with a single channel, in the best

case, 3 votes were lost and in the worst case, 40. All the simulations with Q=1..5 give results. However best results occur with Q= 3 where in the worst case, 24 votes are lost but in the best case, no lost votes are registered.

2) Taking as a reference the simulations with Q=1..5 the worst results are obtained with Q = 1 in all the measured parameters.

Table I show the results obtained for runs of the simulator with Q=1..5, with T=480 slots, N=120 voters and R=1.000.000 repetitions. Similar conclusions were obtained for different values of parameters.

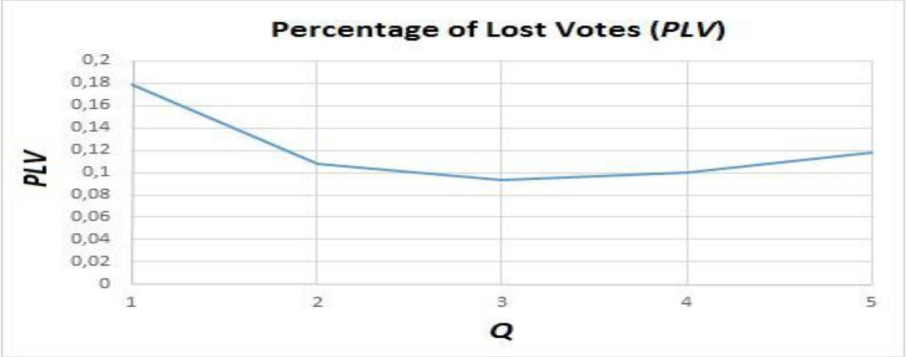

Fig 3: Percentage of Lost Votes

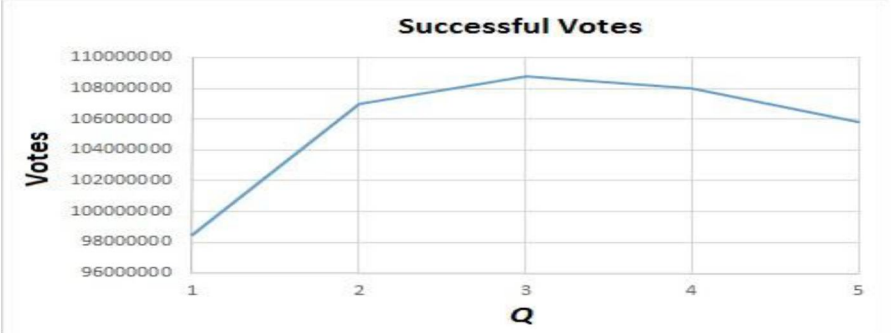

Fig. 4: Simulation Results: *SV*

Q	*SV*	*LV*	*PLV*	*BC*	*WC*
1	98291231	21508769	0,17924	2	40
2	107029947	12970053	0,108084	1	27
3	108763817	11236133	0,0936344	0	24
4	108001192	11998808	0,0999909	1	26
5	105794687	14205313	0,118378	1	30

Table 1: Results of Simulator's Runs

5. CONCLUSIONS

It is considered demonstrated that the proposed technique increases efficiency of storing anonymous data resulting in an advantageous variant respect from the use of a single array. This fact allows setting the security level for the desired value. The empirical results suggest that the behavior is higher in all aspects selected.

The quantity of votes that recovered successfully is optimized dividing T in Q_p slots. The quantity of runs without lost votes is highest when Q_p is used.

The best case (BC) optimizes with the use of Q_p. This means that the run with fewer lost votes occurs when Q_p is used. Similarly this occurs with the variable WC (Worst Case).

Grounds to suggest that the new proposal is a significant improvement for the storage of votes in a scheme Non - Interactive Dining cryptographers. The proposed formulas also allow an easy and accurate administration of the security levels to the system user.

References

[1] Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: "Inicio de la Línea de Investigación: Ingeniería de Software y Defensa Cibernética". XV Workshop de Investigadores en Ciencias de la Computación (WICC). Ps. 769 - 773. ISBN: 9789872817961.

[2] van de Graaf J., Montejano G., García P.: "Optimización de un Protocolo Non-Interactive Dining Cryptographers". Congreso Nacional de Ingeniería Informática / Sistemas de Información CoNaIISI 2013. Córdoba, Argentina.

[3] Chaum D.:"The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". Journal of Cryptology. 1988.

[4] van de Graaf J.: "Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting". In: "Towards Trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:978-3-642-12979-7. 2010.

[5] Chaum D.: "Blind Signatures for Untraceable Payments". Advances in Cryptology Proceedings of Crypto 82 (3):199203.

[6] Fujioka A., Okamoto T., Ohta K.: "A Practical Secret Voting Scheme for Large Scale Elections". AUSCRYPT 1992. LNCS, Vol. 718. Ps. 244 - 251. Springer Heidelberg. 1993.

[7] Flajolet P., Gardy D., Thimonier L.: "Birthday Paradox, Coupon Collectors, Caching Algorithms and Self Organizing Search". Discrete Applied Mathematics 39, ps. 207-223. North-Holland. 1992.

[8] van de Graaf J., Montejano G., García P.: "Optimización de un Esquema Ocuppancy Problem Orientado a E-Voting". Memorias del XV Workshop de Investigadores en Ciencias de la Computación 2013 (WICC 2013). Ps. 749 - 753. ISBN: 9789872817961. 2013.

[9] García P., van de Graaf J., Hevia A., Viola A.: "Beating the Birthday Paradox in Dining Cryptographer Networks". The third International Conference on Cryptology and Information Security in Latin America, Latincrypt 2014. Florianopolis, Brasil. Lecture Notes in Computer Science, Springer (2014).

[10] García P., van de Graaf J., Montejano G., Riesco D, Debnath N., Bast S.: "Storage Optimization for Non Interactive Dining Cryptographers (NIDC)". 12th International Conference on Information Technology: New Generations (ITNG 2015). April 13-15, 2015, Las Vegas, Nevada, USA.